

Forschungsrankings: Artefakte oder Sichtbarkeit der Forschungsarbeit?

Hornbostel, Stefan

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Hornbostel, S. (2006). Forschungsrankings: Artefakte oder Sichtbarkeit der Forschungsarbeit? In M. Stempfhuber (Hrsg.), *In die Zukunft publizieren : Herausforderungen an das Publizieren und die Informationsversorgung in den Wissenschaften ; 11. Kongress der IuK-Initiative der Wissenschaftlichen Fachgesellschaften in Deutschland* (S. 263-277). Bonn: Informationszentrum Sozialwissenschaften. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-25364>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Forschungsrankings: Artefakte oder Sichtbarkeit der Forschungsarbeit?

Stefan Hornbostel

IFQ Institut für Forschungsinformation und Qualitätssicherung, Bonn
Institut für Sozialwissenschaften der Humboldt-Universität zu Berlin

Zusammenfassung

Forschungsrankings gehören inzwischen zum Alltag. Die diesen Rankings zugrunde liegenden Indikatoren, ihre Konstruktion und Angemessenheit werden im Beitrag diskutiert und die Beziehungen zwischen einzelnen Indikatoren und verschiedenen Rankings exemplarisch dargestellt.

1 Einleitung

Was noch vor 10 Jahren von vielen Beobachtern als modisches Strohfeuer abgetan wurde, ist heute fester Bestandteil der Wissenschaftslandschaft geworden: Die Evaluation von Forschung und Lehre. Nicht nur die inzwischen in fast allen Bundesländern in der einen oder anderen Form institutionalisierte Evaluation der Lehre und zunehmend auch der Forschung, sondern ebenso die regelmäßig erscheinenden Rankings zu Forschungs- und Lehrleistungen und last not least die zunehmende Verbreitung von kennzahlgestützten Systemen leistungsorientierter Mittelvergabe deuten darauf hin, dass es sich nicht um eine Modewelle, sondern um eine dauerhafte Veränderung handelt. Eine Veränderung, die zwar bisher nur in geringem Umfang zu einer unmittelbaren Umverteilung finanzieller Ressourcen geführt hat (anders als etwa in Großbritannien), aber zunehmend mittelbare Auswirkungen, sei es in Gestalt der Studienortentscheidung von Studienanfängern, sei es in Form ministerialer Strukturentwicklungsplanungen oder hochschulinterner Profilbildungen zeitigt.

Dabei war zu Beginn dieser Entwicklung in den 1980er Jahren die Forschung das bevorzugte Untersuchungsobjekt, das zunächst aus eher akademischen Interessen mit bibliometrischen Instrumenten vermessen wurde; in den 1990ern rückte jedoch die Evaluation der Lehre deutlich in den Vordergrund und seit einigen Jahren lässt sich beobachten, dass sich das öffentliche Interesse neuerlich auf die Forschung verlagert (vgl. van Raan 2004).

Dieses neue Interesse ist weniger durch aufwändige Entwicklungsarbeit an bibliometrischen Verfahren gekennzeichnet, als vielmehr durch die Anwendung derartiger Verfahren, meist kombiniert mit anderen (nicht-bibliometrischen) Forschungsindikatoren und gekennzeichnet durch eine breite öffentliche Kommunikation der Ergebnisse in Form von Ranglisten: Das CHE zusammen mit der ZEIT und der FOCUS veröffentlichen regelmäßig nationale Forschungsrankings, die DFG informiert ebenfalls regelmäßig mit einem „Förder-Ranking“, der Wissenschaftsrat arbeitet an einer Pilotstudie für ein bundesweites Forschungsrating, die Europäische Kommission legte jüngst den Versuch eines europäischen Forschungsrankings in ausgewählten Fächern vor, das CEST in der Schweiz ermittelt die Mitglieder der internationalen „Champions League“, die Jiao Tong University in Shanghai vermarktet ein „Academic Ranking of World Universities“, und das sind bei weitem nicht alle kursierenden Ranglisten zu nationaler oder internationaler Forschungsperformanz. Zumindest diejenigen Hochschulen, die auf derartigen Ranglisten respektabel platziert sind, nutzen verständlicher Weise die jeweiligen Rankings für ihre Marketingzwecke. Ebenso verständlich ist unter diesen Bedingungen allerdings die Frage danach, ob derartige Rankings Forschungsleistungen tatsächlich angemessen abbilden oder vielmehr künstliche Differenzen produzieren, mit möglicherweise fatalen Fehlsteuerungseffekten in der Folge (vgl. Weingart 2005).

2 Sichtbare und unsichtbare Leistungen?

Es sind allerdings nicht nur die Institutionen des Forschungssystems (Institute, Fachbereiche, Hochschulen), die mit der dauerhaften Bewertung der Forschungsperformanz unter Anforderungsdruck geraten. Auch auf der disziplinären Ebene provozieren die Forschungsrankings eine Auseinandersetzung um die Kriterien angemessener Bewertung. Derartige Auseinandersetzungen waren bisher den methodologischen Debatten vorbehalten und endeten meist in einer mehr oder weniger friedlichen Koexistenz von Paradigmen und Schulen. Über Evaluationen und Rankings kehrt nun die Frage nach einer „disziplinären Gestalt“ gewissermaßen durch die Hintertür zurück. Die Messung von Effektivität – verstanden als Grad der Zielerreichung – setzt voraus, dass disziplinäre Ziele und Standards vorhanden sind, aus denen sich Kriterien der Bewertung ableiten lassen. Effizienzbeurteilungen – also Zweck-Mittel-Relationen – erfordern neben der Erfassung eines Outputs auch eine Inputmessung. Diese Anforderung auf der disziplinären wie auf der institutionellen Ebene, Zielvorstellungen so zu explizieren, dass sie operationalisiert werden können, stößt bei den verschiedenen Disziplinen auf ganz unterschiedliche Voraussetzungen.

Auf der einen Seite stehen die Naturwissenschaften mit einem hohen Internationalisierungsgrad, einer vergleichsweise gut dokumentierten Forschungskom-

munikation, die sich überwiegend auf Fachzeitschriftenbeiträge stützt, einer weitgehend projektförmigen, unter Einsatz von Drittmitteln organisierten Forschung und einer Nachwuchsförderung, die relativ eng in die Forschungsaktivitäten eingebunden ist. Auf der anderen Seite stehen die Geisteswissenschaften, für die diese Bedingungen nicht oder zumindest weitaus weniger zutreffen. Dazwischen lassen sich jene Sozial- und Gesellschaftswissenschaften positionieren, die starke empirische Orientierungen mit geisteswissenschaftlichem Reflektionswissen verbinden. Die Fähigkeit, ein konsistentes Selbstbild zu entwickeln, ist damit bereits recht unterschiedlich ausgeprägt und wird insbesondere in den relativ jungen Disziplinen meist durch interne Ausdifferenzierungsprozesse zusätzlich belastet. Unter solchen Bedingungen entsteht sehr schnell ein öffentlicher Rechtfertigungsdruck, der sich entlang der in anderen Disziplinen eingesetzten Kriterien wissenschaftlicher Leistung entfaltet. Es wird dann schwer zu unterscheiden, ob das angelegte Kriterium ungeeignet ist oder aber durchaus zutreffend ein Problem des Faches beschrieben wird, das im Fach selbst unzureichend thematisiert ist.

Beispiele dafür, wie unter solchen Bedingungen „Artefakte“ entstehen können, lassen sich sicher viele anführen. Hier sei *pars pro toto* auf einen jüngeren Versuch hingewiesen, „die Forschungsleistungen der Soziologie an zehn deutschen Universitäten“ mit Hilfe bibliometrischer Methoden zu erfassen (Litzenberger/Sternberg 2005, S. 174). Die Autoren der Studie, die durchaus die Beschränkungen ihrer Methode sehen, benutzen den Social Sciences Citation Index (SSCI), um die Forschungsleistungen zu vergleichen. Nun ist dies angesichts einer sprachlich stark national orientierten Disziplin und einer extrem geringen Repräsentanz deutscher Soziologiezeitschriften im SSCI ohnehin ein problematisches Vorgehen. Seine eigentliche Brisanz gewinnt es aber weniger aus nationalen Besonderheiten, sondern aus den disziplinären Publikationsgepflogenheiten. In den Sozialwissenschaften spielt – auch international – die Monographie und der Beitrag in Sammelbänden eine sehr große Rolle, wie die kleine Analyse in Abbildung 1 zeigt.

Im SSCI werden jedoch ausschließlich Zeitschriftenbeiträge ausgewertet. Publikations- und Zitationsanalysen auf dieser Basis machen also einen großen Teil der Forschungsarbeit unsichtbar. Welche Effekte dies auf Leistungsvergleiche hat, ist schwer zu sagen, aber es spricht wenig für die Vermutung, dass es „zu keiner systematischen Verzerrung“ kommt (Litzenberger/Sternberg 2005, S. 187). Die ausgeschlossene Literatur gehört jedenfalls sicherlich nicht zum irrelevanten Teil des Forschungsoutputs, sonst wäre sie kaum in den führenden Zeitschriften zitiert worden¹.

1 Die Diversität der Publikationsmedien ist – anders als in den Naturwissenschaften – typisch für geistes- und sozialwissenschaftliche Fächer. Eine ausschließliche Konzentration auf Zeitschriftenartikel blendet daher wichtige Teile der Fachkommunikation aus. Vgl. hierzu

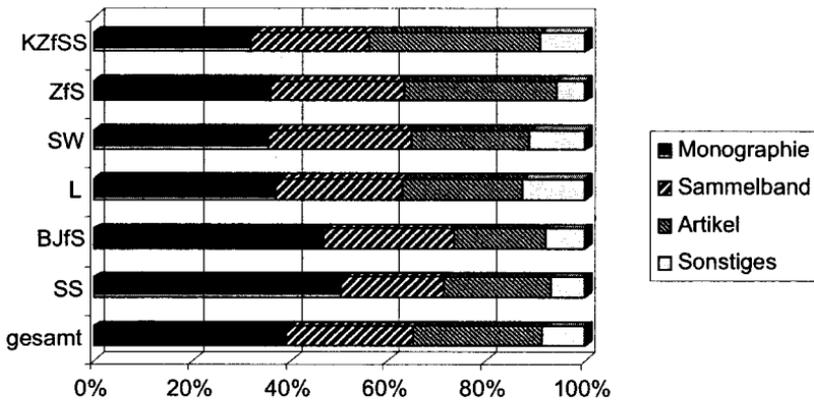


Abb. 1: Wissenschaftliche Quellen, die in sechs deutschen soziologischen Fachzeitschriften des Jahrganges 2002 zitiert wurden – Angaben in Prozent aller Referenzen in der jeweiligen Zeitschrift, Quelle: eigene Erhebungen

3 Artefakte? Natürlich!

Rankings beruhen auf quantitativen Indikatoren und solche Indikatoren sind im ursprünglichen Wortsinne selbstverständlich etwas „künstlich Hergestelltes“. In der sozialwissenschaftlichen Methodenlehre ist mit dem Artefakt allerdings weniger das künstliche oder gar künstlerische der Herstellung gemeint, sondern vielmehr eine Art „Kunstfehler“ im Forschungsprozess, wie am obigen Beispiel dargestellt. Methodisch nicht reflektierte Randbedingungen suggerieren ein Ergebnis, das „prima facie“ als wissenschaftlich solide erscheint und dennoch unzureichend oder gar „falsch“ ist. Artefakte in diesem Sinne sind keine Fälschungen oder offenkundigen Fehler, sie offenbaren sich erst im kritischen Methodendiskurs.

Bei der Konstruktion von Indikatoren wird, in der Regel von einer theoretischen Modellannahme ausgehend, versucht, einen Sachverhalt in seinen wesentlichen Aspekten in einem numerischen Relativ abzubilden. Durch Vergleich

Fortsetzung Fußnote 1

die in der jüngsten Überblicksstudie von Hicks (2004) herausgehobene Bedeutung von vier Kategorien sozialwissenschaftlicher Fachliteratur: Von hoher Relevanz sind neben Veröffentlichungen in internationalen (SSCI) und nationalen Journalen vor allem Monographien und Sammelbände, aber auch die so genannte Transferliteratur (non-scholarly literature), die sich nicht an das direkte Fachpublikum, sondern an Praktiker richtet (vgl. auch Jeanin/Devillard 2005).

erlauben Indikatoren dann auch Bewertungen im Sinne einer mehr oder weniger großen Annäherung an eine Zielvorgabe. Indikatoren sind daher immer begleitet von Objekt- und Instrumenttheorien, um deren Gültigkeit gestritten wird. Sie ähneln insofern durchaus dem wissenschaftlichen Prozedere der Erkenntnisgewinnung, als sie die Randbedingungen ihrer Aussagefähigkeit thematisieren, Vorwissen voraussetzen und als hypothesengeleitete Konstrukte immer wieder an das empirische Wissen angepasst werden müssen.

In diesem Sinne sind Indikatoren hoch artifiziell. Sie konstruieren nämlich gewissermaßen die Welt, die sie abbilden, selbst. Das ist kein Defekt – wie häufig vermutet wird –, sondern eine basale Eigenschaft von Indikatoren. Auffällig wird diese Eigenschaft allerdings nur dort, wo Indikatoren neu entwickelt und implementiert werden. Sind sie einmal zur akzeptierten Routine geworden, wird diese Konstruktionseigenschaft in der Regel übersehen, vergessen oder einfach nicht mehr thematisiert.

Wir haben uns z.B. alle daran gewöhnt, dass wir mit einer Reihe von Indikatoren regelmäßig über den Pulsschlag der Wirtschaft und die Entwicklung gesellschaftlicher Wohlfahrt informiert werden. Sobald das Bruttosozialprodukt steigt oder sinkt, werden Reaktionen ausgelöst. Dabei wird in aller Regel nicht mehr thematisiert, was mit dem Indikator eigentlich gemessen wird, obwohl jeder weiß, dass das Bruttosozialprodukt als Indikator höchst problematisch ist, denn natürlich kann man durch eine stark umweltzerstörende Produktion und die anschließende Reparatur der Schäden das BSP kräftig steigern. Nur: wohlfahrtssteigernd ist dieser Vorgang sicherlich nicht.

Grundsätzlich tritt dieses Problem auch dann auf, wenn Forschungsleistungen nicht aufgrund quantitativer Indikatoren bewertet werden, sondern Beurteilungen durch ein Peer Review-Verfahren erzeugt werden. Auch beim Peer Review entstehen „Artefakte“ im o.g. methodischen Sinne. Der Vorteil des Peer Reviews liegt zweifellos darin, dass Gutachter sehr flexibel die Kontextbedingungen berücksichtigen können, der Preis dafür liegt in einem hohen Grad an subjektiver Färbung des Urteils.

Der wesentliche Unterschied zwischen beiden Verfahren liegt darin, dass dem Gutachtervotum ein expliziter Beurteilungsprozess zugrunde liegt, während Indikatoren massenhaft im Forschungsprozess implizite Urteile (Herausgeberentscheidungen, Mittelbewilligungen etc.) verarbeiten, die zu anderen Zwecken als den mit den Indikatoren verfolgten getroffen wurden.

Weder im Peer Review-Verfahren, noch in einer indikatorenbasierten Bewertung sind es vorrangig wissenschaftsexterne Kriterien, die eine potentielle Produktion von „Artefakten“ vorantreiben, sondern es ist eher eine verstärkte Inanspruchnahme wissenschaftsinterner Kriterien. Die Operationalisierung eines Konzepts von „guter Forschung“ ist es, die die Debatte um die „Artifizialität“ von Forschungsindikatoren immer wieder aufleben lässt.

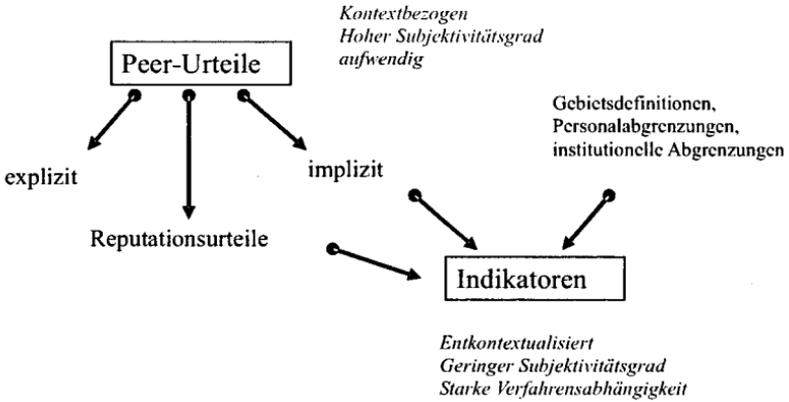


Abb. 2: Bewertungen von Forschungsleistungen

4 Indikatoren

Da die Indikatoren die aus derartigen Operationalisierungsversuchen entstehen, im Großen und Ganzen bekannt sind, beschränke ich mich hier auf einen Überblick (vgl. Hornbostel 1997, 2004, van Raan 2004).

Die erste Gruppe von Indikatoren, bemüht sich darum, Peer-Urteile, die im Förderungs-, Publikations- oder Reputationssystem routinemäßig anfallen, zu quantifizieren. Es kann sich dabei um die Begutachtung von Drittmittelanträgen (Drittmittelindikator), die Manuskriptbegutachtung (Publikationsindikator), Preisverleihungen (z.B. Nobelpreis), Stipendien oder Reputationsurteile (auf Basis von Expertenbefragungen) handeln (vgl. DFG-Förder-Ranking 2003, Sommer 2005). All diese Indikatoren haben spezifische Vor- und Nachteile. Wissenschaftspreise sind meist zu selten oder zu fachspezifisch, um sie für eine Indikatorenbildung zu nutzen, Reputationsurteile weisen nicht nur ein erhebliches time-lag auf, sondern sind auch durch die Quantität des Forschungsoutputs beeinflusst, Drittmittelindikatoren setzen voraus, dass Forschung üblicherweise projektförmig und drittmittelfinanziert organisiert wird, Publikationsindikatoren sagen zunächst etwas über Aktivität, aber nur eingeschränkt etwas über die Qualität der einzelnen Beiträge aus.

Die zweite Gruppe von Indikatoren versucht nicht allein die Peer-Urteile, sondern das Verhalten großer Teile der scientific community im wissenschaftlichen Kommunikationsprozess als Basis zu nutzen. Typischerweise werden in dieser Gruppe nicht nur Publikationen auf der Basis geeigneter Literaturlieferanten erfasst, sondern auch die Reaktionen der community in Gestalt von Zitationen, die Produktionsstrukturen (Kooperationen, Koautorschaften, Netzwerk-

analysen) sowie die Herausbildung thematischer Vernetzungen (Kozitationsanalysen, inhaltsanalytische Untersuchungen). In dieser Indikatorengruppe sind sehr ausgefeilte Analysetechniken entwickelt worden, gleichwohl ist auf der Theorieebene der Status des Zitats nach wie vor umstritten, denn der symbolischen Handlung des Zitierens können sehr unterschiedlich Signifikate zugeordnet werden, so dass die Interpretation von Zitationsindizes – abgesehen von ihrer Gebundenheit an (sub)disziplinäre Gepflogenheiten – noch immer Fragen aufwirft (vgl. Moed 2005).

Eine vollständige Auflistung aller Publikationen ist äußerst schwer zu realisieren, i. d. R. auch nicht erwünscht, weil meist nur ein kleiner Ausschnitt der gesamten Publikationsmenge für relevant gehalten wird (Bradford's Law). Die Abgrenzung „relevanter“ Literatur beginnt mit der Unterscheidung forschungsrelevanter Literatur von anderer (populärwissenschaftlicher, rein dokumentarischer etc.) und endet bei einer relativ scharfen Selektion der Literatur nach Relevanz, Resonanz, Anschlussfähigkeit o.ä.

Für die letztere Variante steht insbesondere der Science Citation Index (SCI) und dessen Derivate, die bereits bei der Auswahl auszuwertender Zeitschriften das Kriterium einer „Mindestsichtbarkeit“ in Form von Zitaten berücksichtigen. Diese fachübergreifende Literaturdatenbank ist also nicht durch Vollständigkeit, sondern (zumindest vom Anspruch her) durch die Auswahl besonders relevanter Nachweise ausgezeichnet. Sie bietet als (bisher²) einzige Datenbank darüber hinaus auch die Möglichkeit, die Zitationshäufigkeit der erfassten Publikationen zu ermitteln. Im Gegensatz dazu erfassen (nationale) Fachdatenbanken ein sehr breites Spektrum (d.h. nicht nur Zeitschriftenartikel, sondern auch Monographien und Sammelbände) der fachlich einschlägigen Forschungsliteratur, meist sogar auch graue Literatur. Diese Datenbanken orientieren sich weitaus stärker an der Vollständigkeit der gesammelten Nachweise, verfügen aber meist weder über klare Auswahlkriterien noch über eine Relevanzeinschätzung der erhobenen Literatur. Schließlich dienen auch fachlich ausgerichtete europäische oder internationale Datenbanken (die allerdings ebenfalls keine Zitate nachweisen) als Informationsquelle.

Besonders relevant wird in Zukunft die Frage werden, wie sich die wachsende Zahl der Open Access-Publikationen angemessen erfassen lässt (vgl. DFG 2005). Insgesamt bietet die Technik des elektronischen Publizierens ein sehr großes Potential für die automatisierte Analyse und Identifikation „relevanter“ Literatur. Allerdings ist bisher das Feld zwischen online-Zeitschriften mit seriösem Review-System, Dokumentenservern und mehr oder weniger privaten

2 Neuerdings bietet auch die Datenbank SCOPUS die Möglichkeit, fachübergreifend nach Zitationen zu suchen. Wie gut sich diese Datenbank für bibliometrische Zwecke eignet, ist bisher aber noch unklar.

websites sehr weitgespannt und die Analysetechniken noch in den Anfängen (vgl. Thelwall/Vaughan/Björneborn 2005).

Eine dritte Gruppe von Indikatoren setzt nicht an der Urteilskompetenz der scientific community an, sondern an den Reaktionen externer Auditorien. Dieser Schritt über die Wissenschaftlergemeinschaft hinaus ist überall dort von Bedeutung, wo Anwendungsbezüge eine wichtige Rolle spielen. Die zugehörigen Indikatoren greifen entweder auf „attestierter Neugigkeit“ (Patentindikatoren), die wirtschaftliche Verwertung (Lizenzentnahmen), die Entwicklung von Kooperationen mit externen Partnern oder auch auf die soziale, kulturelle oder politische Wirkung in einer größeren Öffentlichkeit zurück. Auch hier gilt, dass sowohl Konstruktion wie Interpretation der Indikatoren erheblichen Sachverstand voraussetzen, denn etwa aus Patentanmeldungen lässt sich keineswegs umstandslos auf die Innovationsintensität schließen (vgl. McAleer/Slotje 2005).

Eine letzte – in Rankings häufig benutzte – Gruppe von Indikatoren greift auf die Schnittstelle von Lehre und Forschung, nämlich die Nachwuchsausbildung, zurück. Diese Indikatoren lassen sich nur sehr eingeschränkt als Forschungsindikatoren verstehen. Ob sie mit anderen Indikatoren harmonisieren, hängt wesentlich davon ab, ob die Nachwuchsausbildung eng in das Forschungsprogramm einer Einrichtung integriert ist (wie es häufig in den Naturwissenschaften der Fall ist), oder ob Promotionen weitgehend unabhängig von institutionellen Forschungsprofilen bearbeitet werden (wie häufig in den Geistes- und Sozialwissenschaften der Fall). Davon abgesehen, ist allein die Zahl von Promotionen und Habilitationen wenig aussagefähig, wenn sie nicht um eine Qualitäts- oder Relevanzgewichtung ergänzt wird. Die Tabellen 1 und 2 zeigen sehr deutlich, dass in der Chemie die Zahl der Promotionen pro Professor deutlich positiv mit den per capita-Werten der eingeworbenen Drittmittel und der Publikationen korrelieren und auch zur mittleren Zahl der Zitate pro Artikel eine positive Korrelation besteht. Für die Soziologie hingegen zeigt sich nicht nur kein signifikanter Effekt, sondern sogar eine leicht negative Korrelation.

Die Übersicht in den beiden Tabellen deutet bereits an, dass Forschungsrankings deutlich an Gehalt und Aussagekraft gewinnen, wenn sie sich nicht nur auf einen Indikator stützen, sondern möglichst verschiedene Indikatoren zur Beleuchtung des Konstrukts „Forschungsqualität“ heranziehen. Das ist nicht unbedingt gleichbedeutend mit Multidimensionalität, denn ob sich auf den verschiedenen Indikatoren tatsächlich unterschiedliche Dimensionen forschungsbezogener Aktivitäten abbilden, scheint von Fach zu Fach sehr unterschiedlich zu sein.

Tab. 1: Chemie - Korrelationen zwischen Indikatoren (CHE)**Chemie** - Korrelationen zwischen Indikatoren (CHE)

	Drittm. pro Jahr	DFG- Mittel pro Jahr	Publ. pro Jahr	Zitate pro Jahr	Prom. pro Jahr	Drittm. je Wis- sensch.	Publ. pro Prof.	Zitate pro Papier	Prom. pro Prof.
Reputation	,664**	,532**	,443**	,440**	,635**	,213	,353**	,243*	,447**
Drittmittel pro Jahr		,761**	,572**	,586**	,676**	,640**	,485**	,378**	,500**
DFG-Mittel pro Jahr			,594**	,655**	,568**	,472**	,488**	,455**	,356**
Publikationen pro Jahr				,920**	,716**	,341**	,823**	,411**	,529**
Zitate/Jahr					,721**	,401**	,791**	,694**	,566**
Promotionen pro Jahr						,387**	,508**	,431**	,882**
Drittmittel je Wissensch.							,248*	,351**	,263*
Publikationen pro Professor								,471**	,453**
Zitate/paper									,385**

* Die Korrelation ist auf dem Niveau von 0,05 (1-seitig) signifikant.

** Die Korrelation ist auf dem Niveau von 0,01 (1-seitig) signifikant.

Quelle: CHE Fachbereichserhebung 2002

Tab. 2: Soziologie - Korrelationen zwischen Indikatoren (CHE)**Soziologie** - Korrelationen zwischen Indikatoren (CHE)

	Drittm. insgesamt pro Jahr	DFG- Mittel pro Jahr	Publ. pro Jahr	Prom. pro Jahr	Drittm. je Wissensch. aftler	Publ. je Prof.	Prom. je Prof.
Reputation	,597**	,676**	,501**	,457**	,319*	,357**	,102
Drittmittel ins- gesamt pro Jahr		,740**	,664**	,561**	,534**	,237	,039
DFG-Mittel pro Jahr			,531**	,580**	,327*	,131	,171
Publikationen pro Jahr				,635**	-,003	,615**	,028
Promotionen pro Jahr					-,158	,015	,616**
Drittmittel je Wissenschaftler						,149	-,181
Publikationen je Professor							-,146

* Die Korrelation ist auf dem Niveau von 0,05 (1-seitig) signifikant.

** Die Korrelation ist auf dem Niveau von 0,01 (1-seitig) signifikant.

Quelle: CHE Fachbereichserhebung 2001

5 Fettnäpfchen und Fallstricke

Anders als bei der Evaluation der Lehre werden für die Forschungsevaluationen und darauf basierenden Rankings sehr unterschiedliche Abgrenzungen der Untersuchungseinheiten verwandt. Während in der Lehre ein Studiengang und die ihn verantwortenden Organisationseinheiten maßgeblich sind, mäandert Forschung zwischen organisatorischen und fachlichen Grenzen. Die Wahl einer angemessenen Abgrenzung ist daher weniger eine Frage von richtig und falsch, als vielmehr eine Frage des Erkenntnisinteresses oder der intendierten Anwendung.

Grundsätzlich lassen sich zwei Fragestellungen unterscheiden, die sich auf die Operationalisierung auswirken: Gilt das Interesse vor allen Dingen der Beurteilung organisatorischer Einheiten, dann bildet in der Regel das Personal, das diesen Einheiten zugeordnet ist, die Ausgangsbasis für alle weiteren Analysen. Organisatorische Einheiten wie Lehrstühle, Arbeitsgruppen, Institute, Fachbereiche etc. sind historisch gewachsene Einheiten, die von Hochschule zu Hochschule mehr oder weniger stark in der fachlichen Komposition und Ausrichtung variieren. Insbesondere wenn sich neue Forschungsfelder ausdifferenzieren, kann die Institutionalisierung dieser Gebiete höchst unterschiedlich erfolgen (z.B. kann molekulare Biologie innerhalb der Biologie, der Chemie oder der Medizin beheimatet sein). Umgekehrt beherbergen Fachbereiche u.U. Personal, das von seiner fachlichen Ausrichtung (und entsprechend in der Forschungs- und Publikationsaktivität) auf „fremde“ Forschungsfelder orientiert ist (z.B. Didaktiker in einem naturwissenschaftlichen Fachbereich, die für die Lehrerausbildung zuständig sind). Schließlich werden Kooperationsbeziehungen eingegangen, die sowohl die Organisationsgrenzen als auch die nationalen Grenzen überschreiten.

Der „institutionelle Ansatz“ hat den wesentlichen Vorteil, dass tatsächlich Leistungen einer organisatorisch abgegrenzten Einheit untersucht werden und insofern die Ergebnisse einer Evaluation auch handlungsfähigen Einheiten zugerechnet werden können. Das bedeutet auch, dass sich die „Produktivität“ von organisatorischen Einheiten sehr genau berechnen lässt, da das zugrunde gelegte Personal bekannt ist. Die Nachteile liegen darin, dass u.U. „fachfernes“ Personal in fachbezogenen Datenbanken nicht identifiziert werden kann, dass die Erfassung des akademischen Mittelbaus aufgrund hoher Fluktuation sehr schwierig ist, die Zuordnung von Publikationen teilweise aufwändige Prüfungen der Identität des Autors verlangt (Homonyme) und unter Umständen nur Teile eines Forschungsgebietes erfasst werden (wenn sich ein Forschungsgebiet auf sehr unterschiedliche Institute verteilt). In Deutschland folgt das CHE-Ranking diesem Ansatz.

Der zweite Ansatz geht von einer fachlichen Klassifikation der Drittmittel bzw. der Publikationen aus. Für bibliometrische Analysen bedeutet dies tech-

nisch, dass fachliche Klassifikationen eines Datenbankbetreibers übernommen werden. Meist handelt es sich um die „Subfield-Klassifikation“ des Science Citation Index. Dabei werden nicht Autoren – auch nicht einzelne Zeitschriftenartikel – klassifiziert, sondern ganze Zeitschriften einem Fachgebiet zugeordnet. Das kann konkret bedeuten, dass ein Autor mal als Chemiker, mal als Biologe gezählt wird, je nachdem, wo seine Artikel erschienen sind. Besondere Probleme bereiten dabei die angesehenen multidisziplinären Zeitschriften wie *Science* oder *Nature*, die sich fachlich nicht zuordnen lassen.

Die institutionelle Zuordnung erfolgt dann aufgrund der Adressangabe, die der Autor in seinem Artikel als Korrespondenzadresse angegeben hat. Diese Adressen werfen in der Praxis eine ganze Reihe von Problemen auf, da die Angaben weder vollständig noch standardisiert sind. Das Magazin FOCUS folgt mit seinem Ranking einer Mischung aus Fachklassifikation und Zuordnung nach institutionellen Adressen.

Im Fall des DFG-Rankings dient die interne Zuordnung der Drittmittelanträge zu fachlich kompetenten Fachausschüssen als Klassifikationsgrundlage. Diese Zuordnung ist eine andere als die Klassifikation im Science Citation Index, beide wiederum sind anders als die Klassifikation des Statistischen Bundesamtes. Das DFG-Ranking versucht, durch ausgefeilte Konkordanzen Kompatibilität zwischen diesen abweichenden Klassifikationen herzustellen. Letztlich bleibt aber das jedem Bibliothekar in Gestalt des Thesaurus bekannte Problem, dass Taxonomien nur mehr oder weniger gut auf den (zudem veränderlichen) Gegenstandsbereich passen.

Wie immer man vorgeht, die Abgrenzung der Untersuchungseinheit – mit dem Ziel vergleichbare Einheiten zu definieren – hat immer etwas „Synthetisches“. Problematisch wird dies insbesondere, wenn es um „Effizienz- oder Produktivitätsanalysen“ geht, denn „Sichtbarkeit“ in der Forschung hängt nicht nur von Qualität, sondern auch von einem Größenfaktor ab (je höher der absolute Output an Forschungsergebnissen, desto sichtbarer). Bei der Berechnung von größenunabhängigen pro-Kopf-Angaben wird es sehr wichtig, dem identifizierten „Output“ den zugrunde liegenden personellen „Input“ exakt zuzuordnen. In den CHE-Analysen werden für die bibliometrischen Analysen Namenslisten der Professoren benutzt. In diesem Fall ist also der „Input“ klar definiert und abgegrenzt. Nachteil dieser Strategie ist aber, dass der akademische Mittelbau nicht erfasst wird. Der Mittelbau wird lediglich bei der Kennzahl Drittmittel pro Wissenschaftler berücksichtigt (vgl. Tab. 1 und 2). Auch dies hat allerdings seine Tücken, da die Relation von Professoren und Mittelbau (ohne Drittmittelstellen) sich in Deutschland in den einzelnen Bundesländern sehr unterschiedlich entwickelt hat.

Berücksichtigt man darüber hinaus, dass für die Konstruktion von Indikatoren eine Reihe weiterer Entscheidungen notwendig sind, wie etwa die Zurechnung bei Mehrfachautorschaften, der Umgang mit Selbstziten, die Frage, ob

man nach der „work-done-at“-Methode nur die Publikationen mit der korrekten institutionellen Adresse berücksichtigt oder nach der „current-potential“-Methode alle Publikationen eines Autors, ob man Schwellenwerte oder Gewichtungen benutzt, prozentuale Indizes, Rohwerte oder Abweichungen von Erwartungswerten, dann wird deutlich, dass „Forschungsleistung“ oder „Forschungsqualität“ eine Konstruktion ist, die nicht einfach auf einer Abbildung beruht, sondern auf einem „Herstellungsprozess“. Je klarer innerhalb der jeweiligen scientific communities der Forschungsbegriff und die zugehörigen Qualitätsmerkmale umrissen sind und konsensual getragen werden, desto eher decken sich hinreichend differenzierte Indikatorkonstruktionen mit den Wahrnehmungen der betroffenen Forscher (vgl. van Raan 2005).

Auch die Übereinstimmung unterschiedlicher Forschungsrankings wird je nach Indikator durch die Datengrundlage und die Konstruktion des Indikators mehr oder weniger befriedigend ausfallen. Vergleicht man beispielsweise die Ergebnisse von FOCUS und CHE-Ranking im Fach Psychologie, dann zeigen sich für die Drittmiteleinwerbungen – trotz unterschiedlicher Quellen – mit einer Korrelation von 0,71 noch erhebliche Übereinstimmungen. Für die Zitate je Artikel gilt dies jedoch nicht mehr (Pearson's Corr. 0,29), wie Abbildung 3 zeigt. Neben echten Fehlern (nicht berücksichtigte Publikationen und Zitationen) in beiden Untersuchungen sind es abweichende Konstruktionsentscheidungen, unterschiedliche Personengruppen und nicht ganz deckungsgleiche Untersuchungszeiträume, die zu diesen Abweichungen führen, obwohl die gleiche Datenbasis (SCI) genutzt wurde.

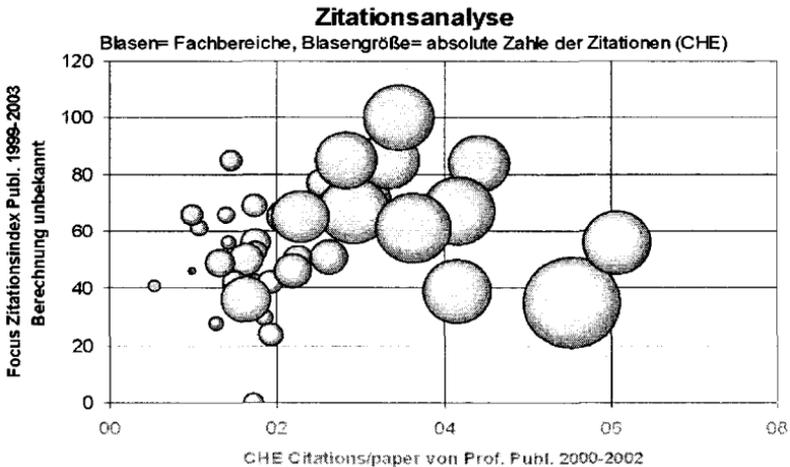


Abb. 3: Zitationsanalyse Fach Psychologie

6 Resümee

Forschungsrankings und -ratings sind einflussreiche Steuerungsinstrumente, wenn sie sorgfältig eingesetzt werden. Umgekehrt haben sie durchaus das Potential fataler Fehlsteuerungen, wenn sie Anreize produzieren, die nicht mit den disziplinären Formen der Wissensakkumulation kompatibel sind. Forschungsindikatoren sind keine Instrumente, die umstandslos in Steuerungsprozessen eingesetzt werden können, sie ähneln eher Diagnoseinstrumenten. Inwieweit sie Forschungsleistungen sichtbar machen können, hängt einerseits an der Solidität ihrer Konstruktion, andererseits daran, wie weit in den jeweiligen Disziplinen Konsens über Qualitätsstandards der Forschung herrscht. Das Grundkonzept von Wissenschaftsindikatoren basiert gerade auf der Überlegung, dass Wissenschaft nur von Wissenschaftlern bewertet werden kann. Solche Bewertungen fallen explizit als Urteile über Forschung oder Forschungsvorhaben an, implizit durch den Umgang mit den Erkenntnisansprüchen anderer. Urteile über die Güte, Nutzbarkeit oder Relevanz von Forschung entstehen also in der Kommunikation über Forschung. Die Urteile, die sich in diesem sozialen Raum bilden, sind nicht mit einer methodologischen Würdigung der Forschungsbefunde zu verwechseln. Die Güte von Forschung in einem methodologischen Sinne besteht nur bedingt aus Akzeptanz und Nutzung. Forschungsindikatoren sind daher Konstruktionen, die bestenfalls das wiedergeben, was Wissenschaftler für gut, wichtig oder nützlich halten.

Die Entwicklung geeigneter Leistungsindikatoren ist ein iterativer Prozess. Erst die Anwendung, empirische Überprüfung und Kritik der Indikatoren führt zu verfeinerten Messinstrumenten. Ein solcher Prozess bedarf allerdings der beständigen Stimulierung und Wissensakkumulation.

Literaturangaben

Deutsche Forschungsgemeinschaft (DFG) (2005): Publikationsstrategien im Wandel? Ergebnisse einer Umfrage zum Publikations- und Rezeptionsverhalten unter besonderer Berücksichtigung von Open Access. Weinheim: Wiley-VCH Verlag.

http://www.dfg.de/dfg_im_profil/zahlen_und_fakten/statistisches_berichtswesen/open_access/download/oa_ber_dt.pdf [Stand: 15.10.05]

Deutsche Forschungsgemeinschaft (DFG) (2003): Förder-Ranking 2003. Institutionen - Regionen - Netzwerke. DFG-Bewilligungen und weitere Basisdaten öffentlich geförderter Forschung. Bonn.

http://www.dfg.de/ranking/download/dfg_foerder_ranking_2003.pdf [Stand: 15.10.05]

- Hicks, D. (2004): The Four Literatures of Social Science. In: Moed, H. F.; Glänzel, W.; Schmoch, U. (Hrsg.): *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems*. Dordrecht: Kluwer Academic. S. 473-496.
- Hornbostel, S. (2004): Leistungsparameter und Ratings in der Forschung. In: *Evaluation – ein Bestandteil des Qualitätsmanagements*. HRK/Beiträge zur Hochschulpolitik 9/2004, S.173-187.
- Hornbostel, S. (1997): *Wissenschaftsindikatoren. Bewertungen in der Wissenschaft*. Opladen: Westdeutscher Verlag.
- Jeannin, P.; Devillard, J. (2005): Implementing relevant disciplinary evaluations in the social sciences. National vs International interactions in scientific communities. In: *Scientometrics*, Vol. 63, Nr. 1, S. 121 – 144.
- Litzenberger, T.; Sternberg, R. (2005): Die Forschungsleistung der Soziologie an zehn deutschen Universitäten. In: *Soziologie*, Jg. 34, Heft 2, S. 74-190.
- McAleer, M.; Slotje, D. (2005): A new measure of innovation: The patent success ratio. In: *Scientometrics*, Vol. 63, Nr. 3, S. 421 – 429.
- Sommer, S. (2005): Bibliometric analysis and private research funding. In: *Scientometrics*, Vol. 62, Nr. 1, S. 165 – 171.
- Moed, H. F. (2005): *Citation Analysis in Research Evaluation*. Dordrecht: Springer.
- Thelwall, M.; Vaughan, L.; Björneborn, L. 2005: Webometrics. In: *Annual Review of Information Science and Technology* 39, S. 81-135.
- Raan A. F.J., van (2005): Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. In: *Scientometrics*, Vol. 62, Nr. 1, S. 133-143.
- Raan A. F.J., van (2004). *Measuring Science*. In: Moed, H. F.; Glänzel, W.; Schmoch, U. (Hrsg.): *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems*. Dordrecht: Kluwer Academic, S. 19-50.
- Weingart, P. (2005): Impact of bibliometrics upon the science system: Inadvertent consequences? In: *Scientometrics*, Vol. 62, Nr. 1, S. 117-131.

Internetquellen

- CEST. „Champions League“: <http://www.cest.ch>
- Shanghai Jiao Tong University. „Academic Ranking of World Universities“: <http://ed.sjtu.edu.cn/ranking.htm>
- Wissenschaftsrat. Forschungsrating: http://www.wissenschaftsrat.de/pilot_start.htm

Angaben zum Autor

Prof. Dr. Stefan Hornbostel
Humboldt-Universität zu Berlin
Philosophische Fakultät III
Institut für Sozialwissenschaften
Unter den Linden 6
10099 Berlin